

Boosting und die Prognose der deutschen Industrieproduktion: Was verrät uns der Blick in die Details?

30

Robert Lehmann und Klaus Wohlrabe

Boosting als eine neuere Methode, große Datensätze für die ökonomische Prognose zu nutzen, wird immer präsenter in der wissenschaftlichen Literatur. Während bisherige Studien oftmals nur die allgemeine Güte von Boosting beurteilen, schaut der vorliegende Artikel in die Prognosedetails des Algorithmus. Konkret gehen wir der Frage nach, welche Indikatoren vom Boosting-Algorithmus zur Vorhersage der deutschen Industrieproduktion im Zeitraum 1996 bis 2014 ausgewählt werden. Im Ergebnis zeigt sich, dass sowohl harte Indikatoren, wie bspw. Auftragseingänge oder Umsätze, als auch Befragungsindikatoren regelmäßig in das Prognosemodell aufgenommen werden. Jedoch enthält der große Datensatz auch Variablen, wie z.B. die Geldmenge, die keine Rolle für die Prognose der deutschen Industrieproduktion spielen.

In der wissenschaftlichen Prognoseliteratur, aber auch in der angewandten Konjunkturanalyse haben sich Methoden, die große Datensätze verarbeiten können (z.B. Faktormodelle oder Kombinationsstrategien), längst etabliert. Neben Bayesianischen Vektorautoregressiven Modellen (BVAR) (vgl. Berg 2015; 2016; Berg und Henzel 2015) drängen Methoden aus dem Feld des Maschinellen Lernens wie bspw. Boosting in den Fokus der aktuellen Forschung. Diese Methoden versuchen, aus einem großen Set von potenziellen Indikatoren möglichst effizient die relevanten herauszufiltern. Selbiges mit Standardmethoden zu versuchen, wäre äußerst zeit- und rechenintensiv.

Bis dato evaluiert die relativ kleine Boosting-Literatur fast ausschließlich die Güte dieser Modellklasse. Wesentliche Beiträge zur Literatur liefern Buchen und Wohlrabe (2011; 2014) für die Industrieproduktion und Pierdzioch et al. (2015; 2016) für Renditen auf Gold und Silber. Neben der spannenden Frage nach der allgemeinen Prognosegüte von Boosting kann der Blick in die Details der Prognosemodelle wichtige Erkenntnisse liefern. In diesem Beitrag fragen wir daher: Gibt es Indikatoren, die durch den Algorithmus regelmäßig für die Prognose herangezogen werden? Für die Vereinigten Staaten von Amerika gibt es dazu bereits eine Studie (vgl. Kim und Swanson 2014). Für Deutschland hingegen fehlt eine entsprechende Analyse.¹ Die Ergebnisse verdeutlichen, dass eine Reihe von Daten der amtlichen Statistik wie bspw. Umsätze und eine kleine Anzahl an Be-

fragungsindikatoren vom ifo Institut wiederkehrend vom Algorithmus zur Prognose herangezogen werden.

Der Boosting-Algorithmus

Die grundlegende Idee von Boosting besteht in der iterativen Schätzung einer unbekannt linearen oder nichtlinearen Funktion. Dieser iterative Prozess geschieht solange, bis gewisse Kriterien erfüllt sind, so dass der Algorithmus letztendlich konvergiert. An dieser Stelle werden alle technischen Details, mit Verweis auf Lehmann und Wohlrabe (2016), ausgeblendet. Für große Datensätze, bei denen die Anzahl der Indikatoren die Zeitdimension übersteigt, muss zunächst die Komplexität durch Vorauswahl der Variablen reduziert werden (vgl. Bühlmann und Yu 2003). Dazu schätzt der hier verwendete Boosting-Ansatz ein allgemeines additives Modell, das wir in Form des sogenannten autoregressive distributed lag (ADL-)Modells spezifiziert haben:

$$E(y_t | \mathbf{z}_t, \delta) := F(\mathbf{z}_t, \delta) = \alpha + \beta_1 y_{t-h} + \sum_{n=1}^N \gamma_n x_{t-h}^n.$$

Die Zielvariable y_t wird erklärt durch eine Konstante α , ihren um h -Perioden verzögerten Wert y_{t-h} und die Verzögerungen exogener Indikatoren x_{t-h}^n . Insgesamt nutzen wir N exogene Indikatoren, die dem Algorithmus zur Verfügung stehen. Bei β_1 und γ_n handelt es sich um die zu schätzenden Koeffizienten. Alle Indikatoren, die nicht vom Algorithmus ausgewählt werden, erhalten einen Wert des Koeffizienten gleich null.

¹ Der vorliegende Aufsatz basiert auf der Publikation von Lehmann und Wohlrabe (2016).

Zur Entscheidung, ob ein Indikator in das Modell aufgenommen wird, nutzen wir eine quadratische Verlustfunktion mit folgender Form:

$$L(y_t, F(\mathbf{z}_t, \delta)) = \frac{1}{2}(y_t - F(\mathbf{z}_t, \delta))^2.$$

Der Algorithmus wählt in jedem Schritt einen Indikator aus, der die Residuenquadratsumme minimiert und damit die beste Anpassung an die Zielzeitreihe erzielt. Es kann aber durchaus vorkommen, dass ein Indikator mehrfach vom Algorithmus in das Modell selektiert wird. Zur Schätzung der Parameter verwenden wir in diesem Fall die Methode der Kleinsten Quadrate (im Zusammenhang mit Boosting wird die Schätzmethode auch als »Lerner« bezeichnet). Der Algorithmus gibt letztendlich ein Modell aus, das eine optimale Anzahl an Indikatoren M^* berücksichtigt. Jedoch werden die einzelnen Indikatoren nicht einfach additiv hintereinander geschaltet, sondern mit dem »Regulierungsparameter« ν multipliziert:

$$\hat{F}(\mathbf{z}_t, \hat{\delta}^{[M^*]}) = \sum_{m=0}^{M^*} \nu \hat{f}_m(\mathbf{z}_t, \hat{\theta}^{[m]}).$$

Der »Regulierungsparameter« kann dabei beliebige Werte zwischen 0 und 1 annehmen. Wir folgen der einschlägigen Literatur und nehmen einen Wert von 0,1 an. Der wesentliche Grund für die Einführung des »Regulierungsparameters« ist die Reduktion der Varianz des »Lerners« (vgl. Friedman 2001). Damit soll die Güte der Prognosen verbessert werden.

Datensatz

Als Zielzeitreihe (y_t) verwenden wir die deutsche Industrieproduktion. Um Stationarität der Industrieproduktion zu erreichen, transformieren wir die Zeitreihe in Wachstumsraten zum Vorjahr. Die Daten liegen in monatlicher Frequenz im Zeitraum Januar 1996 bis Dezember 2014 vor. Zur Prognose der Industrieproduktion nutzen wir einen Datensatz, der aus 175 Indikatoren besteht, die wir grob in fünf Gruppen zusammenfassen können: (i) makroökonomische Variablen, (ii) Finanzvariablen, (iii) Preise, (iv) Befragungsdaten und (v) internationale Indikatoren. In der ersten Gruppe sind Indikatoren wie bspw. Umsätze oder Auftragseingänge subsumiert. Gruppe 2 enthält Indikatoren wie bspw. den Deutschen Aktienindex (DAX) oder die Geldmenge M3. Der Verbraucherpreisindex oder der Ölpreis fallen in Gruppe 3. Indikatoren, die aus wichtigen Befragungen gewonnen werden (z.B. ifo Institut oder ZEW), finden sich in der vierten Gruppe. Den Abschluss bilden internationale Indikatoren wie bspw. die amerikanische Industrieproduktion oder die Composite Leading Indicators der Organisation für wirtschaftliche

Zusammenarbeit und Entwicklung (OECD) in Gruppe 5. Aufgrund der Offenheit der deutschen Volkswirtschaft erwarten wir eine gewisse Güte der internationalen Indikatoren zur Vorhersage der Industrieproduktion. Aus diesem Grund umfasst die letzte Indikatorengruppe Variablen wichtiger deutscher Handelspartner wie Frankreich oder die Volksrepublik China. Alle Indikatoren sind saisonbereinigt und transformiert in erste Differenzen oder Jahreswachstumsraten, falls notwendig.

Prognoseansatz

Die Prognosehorizonte der Jahreswachstumsrate der deutschen Industrieproduktion sind $h = 1, 3, 6, 12$ Monate. Somit erstellen wir eine Prognose, die maximal ein Jahr in die Zukunft reicht. Alle Prognosen werden direkt erzeugt, d.h. ohne die Vorhersage der unbekanntenen Indikatoren im Prognosezeitraum. Die erste Stützperiode zum Schätzen der Parameter erstreckt sich von Januar 1996 bis Dezember 2004 und wird in jedem Iterationsschritt um einen Monat verlängert (expandierendes Fenster). Somit erzeugen wir die erste Prognose der Industrieproduktion für Januar 2005. Als Gütemaß verwenden wir die Wurzel aus dem mittleren quadratischen Prognosefehler; den Vergleichsmaßstab bildet ein autoregressiver Prozess erster Ordnung. Dieser Prozess ergibt sich, wenn beim o.g. ADL-Modell neben den Verzögerungen der Zielzeitreihe kein zusätzlicher Indikator integriert wird. Alle weiteren Parameter bzw. die Kriterien zur Modellbestimmung sind, wie von Buchen und Wohlrabe (2014) vorgeschlagen, optimal gewählt.

Ergebnisse

Bevor wir in die Details der Ergebnisse einsteigen, präsentieren wir zunächst vier allgemeine Resultate. Erstens generiert Boosting für alle vier Prognosehorizonte geringere Prognosefehler als das Benchmark-Modell, was die Resultate aus Buchen und Wohlrabe (2014) bestätigt. Die Relationen der Prognosefehler von ADL-Modell zu Benchmark betragen 0,881, 0,809, 0,802 und 0,950 für $h = 1, 3, 6, 12$.² Zweitens variiert die Zusammensetzung der zehn besten Indikatoren mit dem Prognosehorizont. Jedoch gibt es Variablen, die eine hohe Prognosegüte für alle Horizonte aufweisen. Als drittes zentrales Ergebnis finden wir, dass Indikatoren wie bspw. die Geldmenge im Evaluationszeitraum nie vom Algorithmus ausgewählt werden.³ Makroökonomische Indikatoren und Befragungsergebnisse liefern die besten Prognosen im Rahmen unseres Boosting-Ansatzes, was zugleich das vierte wesentliche Ergebnis unserer Studie ist.

² Werte kleiner als 1 signalisieren eine höhere Prognosegüte des Boosting-Modells gegenüber dem Benchmark.

³ In der Arbeitspapierfassung dieses Beitrages sind im Appendix alle Indikatoren und die Anzahl, wie häufig diese ausgewählt werden, aufgelistet (vgl. Lehmann und Wohlrabe 2015).

Einen Blick in die Boosting-Details bietet Abbildung 1. Für jeden Prognosehorizont präsentiert die Abbildung die fünf am häufigsten vom Algorithmus gewählten Indikatoren über die Zeit. Die beiden oberen Graphiken zeigen die kürzeren Prognosehorizonte; im unteren Teil sind die beiden längsten Prognosehorizonte abgebildet. Die Interpretation aller Graphiken ist deckungsgleich. Auf den x-Achsen ist jeweils der Prognosezeitraum (Januar 2005 bis Dezember 2014, $T = 120$) abgetragen. Die y-Achsen zeigen die kumulierten relativen Häufigkeiten wie oft ein Indikator gewählt wurde. Demzufolge liegt der Definitionsbereich der y-Achsen zwischen 0% (Indikator nie ausgewählt) und 100% (Indikator zu jedem Zeitpunkt ins Modell aufgenommen). Sofern ein Indikator zum Zeitpunkt t vom Modell gewählt wird, steigt seine relative Häufigkeit um $1/120$ bzw. 0,083 Prozentpunkte. Die gesamte relative Häufigkeit eines Indikators ist letztendlich die Summe der Zeitpunkte, an denen der Indikator ausgewählt wird, geteilt durch 120. Sofern ein Indikator zu jedem Zeitpunkt ins Modell selektiert wird, würde die daraus resultierende Linie in den Graphiken einen Winkel von 45° annehmen. In Phasen, in denen der Indikator nicht gewählt wird, verläuft die entsprechende Linie horizontal.

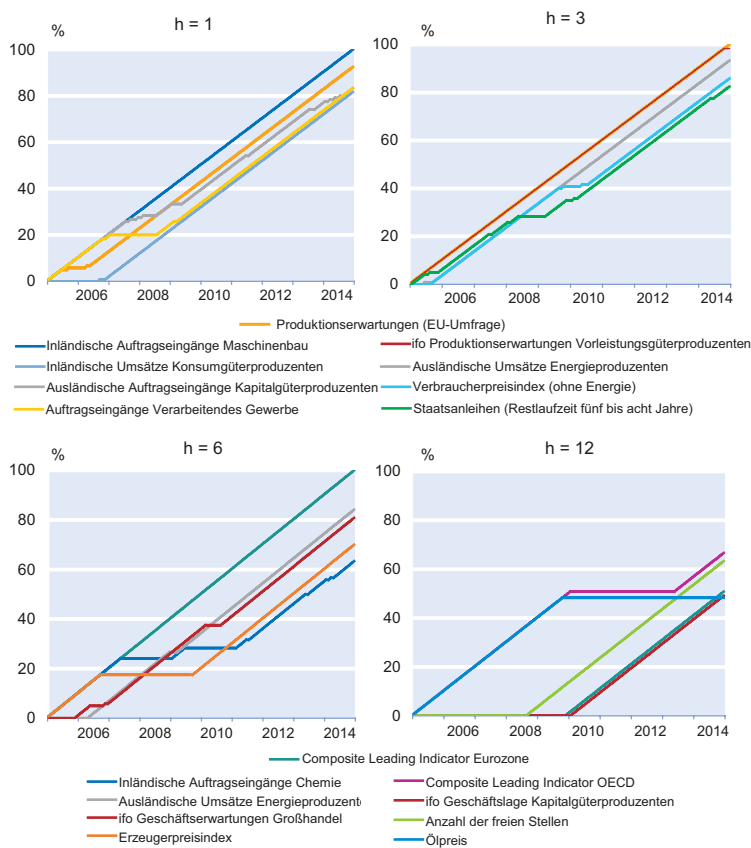
Der beste Indikator in der kurzen Frist ($h = 1$) sind die inländischen Auftragseingänge im Maschinenbau. Zu jedem Zeitpunkt im Prognosezeitraum ist diese Variable durch den Algorithmus in das Modell gewählt worden, was eindeutig durch die 45° -Linie angezeigt wird. Der am zweithäufigsten gewählte Indikator sind die Produktionserwartungen aus der monatlichen Unternehmensbefragung der Europäischen Union. Neben diesen beiden Indikatoren sind besonders die Auftragseingänge der gesamten deutschen Industrie sowie jene aus dem Ausland für die deutschen Kapitalgüterproduzenten. Unter den Top 10 findet sich auch ein Indikator des ifo Instituts: die Produktionserwartungen der Konsumgüterproduzenten.

Die besondere Stellung der ifo-Indikatoren wird deutlich bei Betrachtung des Prognosehorizonts $h = 3$. Neben den ifo Produktionserwartungen der deutschen Vorleistungsgüterproduzenten, dem zweithäufigsten vom Algorithmus gewählten Indikator, finden sich weiterhin drei ifo-Befragungsindikatoren (ifo Geschäftserwartungen für den Einzelhandel, das Verarbeitende Gewerbe und die Konsumgüterproduzenten) unter den Top 10. Eine wesentliche Stellung für $h = 3$ nehmen die EU-Produktionserwartungen ein, die der am häufigsten gewählte Indikator für diesen Prognosezeitraum sind. Neben den genannten Indikatoren werden insbesondere die ausländischen Umsätze der Energieproduzenten, der Verbraucherpreisindex und Staatsanleihen mit einer Restlaufzeit von fünf bis acht Jahren regelmäßig ins Modell selektiert.

Für den Prognosehorizont $h = 6$ kristallisiert sich der Composite Leading Indicator (CLI) für die Eurozone als der beste Indikator heraus, gefolgt von den ausländischen Umsätzen der Energieproduzenten. Erneut befindet sich ein ifo-Indikator, die Geschäftserwartungen im deutschen Großhandel, unter den fünf am häufigsten ausgewählten Variablen. Ein Blick auf die Top 10 beweist erneut die herausragende Stellung der Befragung des ifo Instituts: weitere drei Befragungsindikatoren werden regelmäßig in das Modell aufgenommen, darunter erneut die Produktionserwartungen der Vorleistungsgüterproduzenten.

In der langen Prognosefrist ($h = 12$) sind es besonders die CLI für die OECD und die Eurozone, die wichtige Signale zur Vorhersage der Industrieproduktion liefern. Zusätzlich finden sich unter den fünf am häufigsten gewählten Indikatoren auch die freien Stellen, die ifo Geschäftslage der Kapitalgüterproduzenten sowie der Ölpreis.

Abb. 1
Die fünf am häufigsten vom Algorithmus gewählten Indikatoren nach Prognosehorizont



Quelle: Berechnungen des ifo Instituts.

Schlussfolgerungen

Der vorliegende Artikel unterstreicht, dass Boosting durchaus einen wesentlichen Beitrag zur Verbesserung der Prognose der Industrieproduktion liefern kann. Besonders spannend ist der Blick in die Details: Sowohl harte Indikatoren wie die Auftragseingänge als auch Befragungsindikatoren stellen Variablen dar, die regelmäßig vom Algorithmus in das Prognosemodell gewählt werden. Jedoch finden wir ebenfalls, dass eine Vielzahl an Indikatoren wie bspw. die Geldmenge keine Rolle spielt.

Was folgt letztendlich aus den Erkenntnissen der vorliegenden Studie? Wir können zwei wesentliche Aspekte identifizieren. Erstens folgen aus diesem Artikel Hinweise für die weitere Forschung. In der Regel fokussiert sich die Prognoseliteratur auf Größen wie das Bruttoinlandsprodukt oder die Industrieproduktion. Jedoch erwarten wir spannende Erkenntnisse für andere Größen wie bspw. die Arbeitslosenquote oder andere Arbeitsmarktvariablen. Darüber hinaus können weitere Studien eine Vielzahl an Methoden für große Datensätze (z.B. Prognosekombinationen, Faktormodelle und BVARs) miteinander vergleichen und ggf. Vorteile in bestimmten Situationen herausarbeiten. Der zweite Aspekt ergibt sich für die Prognosepraxis. Basierend auf unseren Erkenntnissen, kann Boosting als durchaus kompetitiv für die praktische Prognose angesehen werden. Somit wäre ein »Tauglichkeitstest« in der Praxis ein erster Schritt, Boosting in der Prognosepraxis zu etablieren.

Literatur

Berg, T.O. (2015), »Forecast Accuracy of a BVAR under Alternative Specifications of the Zero Lower Bound«, Ifo Working Paper Nr. 203.

Berg, T.O. (2016), »Multivariate Forecasting with BVARs and DSGE Models«, *Journal of Forecasting*, im Erscheinen.

Berg, T.O. und S.R. Henzel (2015), »Point and Density Forecasts for the Euro Area Using Bayesian VARs«, *International Journal of Forecasting* 31(4), 1067–1095.

Buchen, T. und K. Wohlrabe (2011), »Forecasting with many predictors: Is boosting a viable alternative?«, *Economics Letters* 113(1), 16–18.

Buchen, T. und K. Wohlrabe (2014), »Assessing the Macroeconomic Forecasting Performance of Boosting – Evidence for the United States, the Euro Area, and Germany«, *Journal of Forecasting* 33(4), 231–242.

Friedman, J.H. (2001), »Greedy Function Approximation: A Gradient Boosting Machine«, *The Annals of Statistics* 29(5), 1189–1232.

Kim, H. H. und N. R. Swanson (2014), »Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence«, *Journal of Econometrics* 178(2), 352–367.

Lehmann, R. und K. Wohlrabe (2015), »Looking into the Black Box of Boosting: The Case of Germany«, CESifo Working Paper Nr. 5686, verfügbar unter: https://ideas.repec.org/p/ces/ceswps/_5686.html.

Lehmann, R. und K. Wohlrabe (2016), »Looking into the black box of boosting: The case of Germany«, *Applied Economics Letters*, im Erscheinen.

Pierdzioch, C., M. Risse und S. Rohloff (2015), »Forecasting gold-price fluctuations: a real-time boosting approach«, *Applied Economics Letters* 22(1), 46–50.

Pierdzioch, C., M. Risse und S. Rohloff (2016), »A boosting approach to forecasting gold and silver returns: economic and statistical forecast evaluation«, *Applied Economics Letters*, im Erscheinen.